

In []:



```
from google.colab import drive
drive.mount('/content/gdrive')
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call drive.mount("/content/gdrive", force_remount=True).

In []:



```
cd gdrive/MyDrive/PS/
```

[Errno 2] No such file or directory: 'gdrive/MyDrive/PS/'
/content/gdrive/MyDrive/PS

In []:



```
import pandas as pd
import numpy as np
import math
```

In []:



```
df=pd.read_csv('US_Confirmed_daily.csv')
```

In []:



```
#df=df.set_index('Date')
df['Date']=pd.to_datetime(df['Date'])
df
per = df.Date.dt.to_period("M")
g = df.groupby(per)
df_conf=g.sum()
df_conf
```

Out[150]:

Confirmed	
Date	
2020-01	8
2020-02	31
2020-03	213096
2020-04	883270
2020-05	704186
2020-06	867713
2020-07	1918996
2020-08	1433328
2020-09	1201058
2020-10	1896322
2020-11	4414607
2020-12	6319937
2021-01	6033109
2021-02	2295725
2021-03	1766512
2021-04	131255

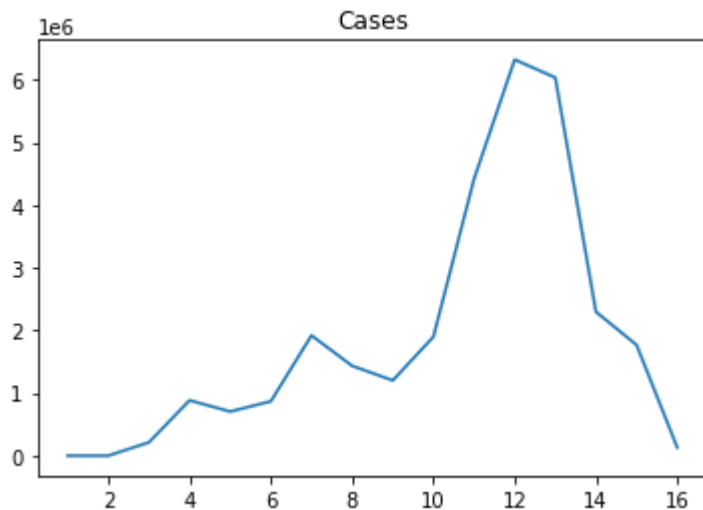
In []:

```
import matplotlib.pyplot as plt
Xaxis= np.linspace(1, 17,16, endpoint=False)
Yaxis=df_conf['Confirmed'].to_numpy()
plt.plot(Xaxis,Yaxis)
plt.title("Cases")
plt.show()
df_imp=pd.read_csv('Imports.csv')
Y_imp=df_imp['Imports'].to_numpy()
Y_imp=Y_imp.astype(np.float)

# X_imp= np.linspace(1, 16,15, endpoint=False)
# plt.plot(X_imp,Y_imp)
# plt.title("Imports")
# plt.show()

df_exp=pd.read_csv('Exports.csv')
Y_exp=df_exp['Exports'].to_numpy()
Y_exp=Y_exp.astype(np.float)

X_exp= np.linspace(1, 16,15, endpoint=False)
plt.plot(X_exp,Y_exp)
plt.title("Exports")
plt.show()
```



Exports

In []:

```
def pearson_correlation(X,Y):
    rows=min(X.shape[0],Y.shape[0])
    X=X[:rows]
    Y=Y[:rows]
    X_mean=np.mean(X)
    Y_mean=np.mean(Y)
    print(X_mean,Y_mean)
    sum=0
    sum_X=0
    sum_Y=0
    for i in range(rows):
        sum+=(X[i]-X_mean)*(Y[i]-Y_mean)
        sum_X+=(X[i]-X_mean)*(X[i]-X_mean)
        sum_Y+=(Y[i]-Y_mean)*(Y[i]-Y_mean)
    corr=sum/(math.sqrt(sum_X)*math.sqrt(sum_Y))
    return corr
conf_np=df_conf['Confirmed'].to_numpy()[:5]
exp_np=df_exp['Exports'].to_numpy()[:5]
print("Pearson Correlation Coefficient in first 5 months=",pearson_correlation(conf_np,exp_
```

360118.2 178658.8

Pearson Correlation Coefficient in first 5 months= -0.9736220086837345

In []:

```
conf_np=df_conf['Confirmed'].to_numpy()[5:]
exp_np=df_exp['Exports'].to_numpy()[5:]
print("Pearson Correlation Coefficient after first 5 months=",pearson_correlation(conf_np,e
```

2814730.7 181381.1

Pearson Correlation Coefficient after first 5 months= 0.5387566444268533

Results of Inference 1:

Hypothesis:

Rise in Covid Cases in US Negatively Impacted the export of goods for the first five months.

Alternate Hypothesis:

Rise in Covid Cases in US Positively Impacted the export of goods for the first five months.

Results:

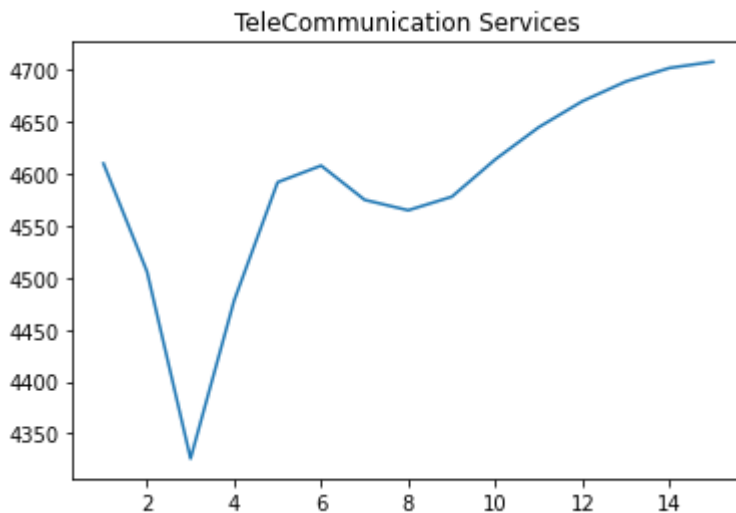
Since the pearson correlation coefficient is closer to -1 we can say that our hypothesis is true and they are infact negatively correlated.

After the first 5 months they are only moderately correlated, with a value of +0.53, which indicates the resumption of services and relaxtion of rules for the exports of goods.

In []:

```
df_Tele=pd.read_csv('Tele.csv')
Y_ins=df_Tele['Tele'].to_numpy()
Y_ins=Y_ins.astype(np.float)

X_ins= np.linspace(1, 16,15, endpoint=False)
plt.plot(X_ins,Y_ins)
plt.title("TeleCommunication Services")
plt.show()
```



In []:

```
Tele_first=df_Tele['Tele'].to_numpy()[:6]
Tele_last=df_Tele['Tele'].to_numpy()[6:12]
Var_first = ((Tele_first - Tele_first.mean()) ** 2).sum()/(len(Tele_first) - 1)
Var_last = ((Tele_last - Tele_last.mean()) ** 2).sum()/(len(Tele_last) - 1)
print(Var_first,Var_last)
pool_std = (Var_first / len(Tele_first) + Var_last / len(Tele_last)) ** 0.5

t = abs((Tele_first.mean() - Tele_last.mean()) / pool_std)

print("T-value =",t)

reject_hypo = (t >= 2.0423)

print("H0: mean(mar)=mean(feb), Reject:\n", reject_hypo)
```

```
12124.8 1817.3666666666663
T-value = 1.8220901273838341
H0: mean(mar)=mean(feb), Reject:
False
```

Results of Inference 2

Hypothesis:

Telecommunications, Computer, and Information Services exports remained the same, that is covid has no impact on this sector.

Alternate Hypothesis:

Telecommunications, Computer, and Information Services exports were impacted due to Covid.

Procedure:

The covid cases during the period Jan-2020 to Jun-2020(1st half of year) are comparatively very less when compared to the period between Jul-2020 to Dec-2020(2nd half of year).

We would be using the unpaired 2-sample t-test to compare the means in both the 1st half and 2nd half of the year.

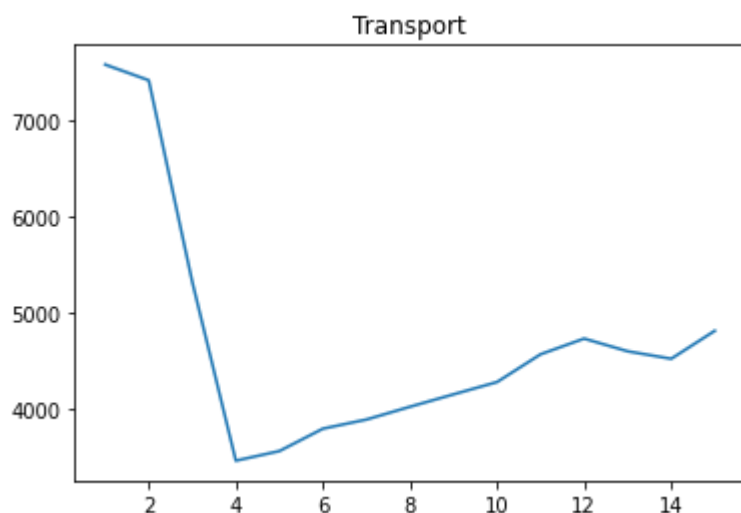
Result:

We got $t=1.822 < 2.04$, hence we accept the hypothesis. So, covid has no impact on the Telecommunications, Computer, and Information Services exports sector.

In []:

```
df_Tns=pd.read_csv('Transport.csv')
Y_Tns=df_Tns['Transport'].to_numpy()
Y_Tns=Y_Tns.astype(np.float)

X_Tns= np.linspace(1, 16,15, endpoint=False)
plt.plot(X_Tns,Y_Tns)
plt.title("Transport")
plt.show()
```



In []:



```
def get_beta(x,y):
    x_mean = np.mean(x)
    y_mean = np.mean(y)
    xy = x * y
    xy_mean = np.sum(xy)
    x_sq = x * x
    x_sq_mean = np.sum(x_sq)
    n = x.size
    beta1 = (xy_mean - (n * x_mean * y_mean)) / (x_sq_mean - (n * x_mean * x_mean))
    beta0 = y_mean - (x_mean * beta1)
    return beta0,beta1
```

```
X=df_conf['Confirmed'].to_numpy()[:]
Y=df_Tns['Travel'].to_numpy()
b0,b1=get_beta(X,Y)
```

In []:



```
from sklearn import metrics as met
import sklearn.metrics as sm

X=df_conf['Confirmed'].to_numpy()[:-1]
df_Tns=pd.read_csv('Travel.csv')[:]
Y_Tns=df_Tns['Travel'].to_numpy()
Y=df_Tns['Travel'].to_numpy()

b0,b1=get_beta(X[:12],Y[:12])
Y_pred=b0+b1*X

print("Row|Predicted|Actual")
for i in range(X.shape[0]):
    print(i,"|",Y_pred[i],"|",Y[i])
print("Explained Variance =", round(sm.explained_variance_score(Y, Y_pred), 2))
```

```
Row|Predicted|Actual
0 | 11149 | 7145
1 | 9250 | 9438
2 | 3899 | 3048
3 | 173 | 891
4 | 610 | 755
5 | 1675 | 741
6 | 1251 | 657
7 | 556 | 1336
8 | 251 | 1628
9 | 768 | 1817
10 | 3028 | 1500
11 | 2971 | 2836
12 | 2872 | 2593
13 | 2947 | 2318
14 | 3883 | 2466
Explained Variance = 0.71
```

Results of Inference 3

Hypothesis:

Travel sector is linearly dependent on covid cases.

Alternate Hypothesis:

Travel sector is independent on covid cases.

Procedure:

Using linear regression we predicted the values for travel for the months of Jan2020-Mar2021. Used the explained variance score metric to calculate the error between actual and predicted values.

Result:

We got Explained variance score = $0.71 > 0.5$, hence we say travel sector is dependent on Covid cases.

Type *Markdown* and LaTeX: α^2